Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Procedia CIRP 23 (2014) 178 – 183

www.elsevier.com/locate/procedia

Conference on Assembly Technologies and Systems

Data Mining-supported Generation of Assembly Process Plans

R. Wallis[a]*, O. Erohin[b], R. Klinkenberg[c], J. Deuse[b], F. Stromberger[a]

[a]*Daimler AG, Hanns-Martin-Schleyer-Str. 21-57, 68305 Mannheim, Germany*
[b]*TU Dortmund University, Institute of Production Systems, Leonhard-Euler-Str. 5, 44227 Dortmund, Germany*
[c]*RapidMiner GmbH, Stockumer Str. 475, 44227 Dortmund, Germany*

* Corresponding author. Tel.: +49-621-393-7734; fax: +49-711-3052115938. *E-mail address:* regina.wallis@daimler.com

Abstract

The application and functional scope of digital assembly planning tools have been permanently increasing in order to deal with product and process complexity. Consequently a large amount of assembly-related data is stored in different systems alongside the product emergence process. By means of data mining techniques an intelligent utilization of this data can be accomplished for future assembly planning. This paper presents an approach for data mining-supported generation of assembly process plans to enhance planning efficiency. The approach is based on the classification and clustering of both product and process data as well as on the identification of their correlations.

Selection and peer-review under responsibility of the International Scientific Committee of 5th CATS 2014 in the person of the Conference Chair Prof. Dr. Matthias Putz matthias.putz@iwu.fraunhofer.de

Keywords: assembly; digital process planning; data mining; variant management; planning support

1. Introduction

Manufacturing companies are facing the challenge of developing and producing a continuously rising number of product variants in shorter periods of time in order to be competitive. Especially in assembly planning the resulting process complexity becomes apparent [1] and is difficult to keep under control, as all product and process variants have to be kept at the same time. Manufacturing industry and the automotive industry as a pioneer, has consequently been concentrating on the application of digital manufacturing systems in order to counteract these challenges as they allow to support simultaneous engineering processes and to cope with the multitude of product variants. As a result, big digital databases concerning product and assembly process planning are available nowadays. These databases possess the potential for an improvement of the assembly planning process: By identifying correlations and recurrent patterns, data mining techniques provide an opportunity to discover tacit planning knowledge in these databases, which subsequently can be reintegrated into new assembly planning workflows to enhance planning efficiency and facilitate decision making.

This paper presents a novel approach for the data mining-supported generation of assembly process plans based on the data compiled during the product emergence process. Therefore Section 2 reviews the current state of the art in assembly planning, classification and clustering methods as well as their integration in the procedure of knowledge discovery in databases (KDD). Section 3 subsequently summarizes the challenges and requirements for the improvement of modern assembly process planning. Sections 4 and 5 present the concept of data mining-supported assembly planning and demonstrate first application results in the automotive industry. Section 6 summarizes the results and identifies future research activities.

2. Knowledge discovery based on assembly-related data

The intensified application of digital manufacturing and product data management (PDM) systems results in a large amount of data allowing to document the product emergence

Selection and peer-review under responsibility of the International Scientific Committee of 5th CATS 2014 in the person of the Conference Chair Prof. Dr. Matthias Putz matthias.putz@iwu.fraunhofer.de
doi:10.1016/j.procir.2014.10.095

process thoroughly. The utilization of this product and process data by application of data mining techniques provides a basis for the discovery of the hidden assembly relevant knowledge.

2.1. Product and process data in assembly planning

Product data represents a relevant input dataset for the assembly planning process and is typically stored in PDM systems, which have been developed in the context of Computer Aided Design to support product development and construction. They provide the central storage and management of product-related data [2]. The enhancements of PDM systems are nowadays represented by Product Lifecycle Management solutions, which focus on the system application alongside the entire product lifecycle including process planning, production and after sales management.

First support functions for assembly planning have been developed in the context of the Computer Aided Process Planning systems, which are able to generate work plans based on the product description. The required planning knowledge is provided in an organized and formalized way, e.g. in the form of “if-then-else”-rules [3] or in form of decision trees [4]. However, with an increasing number of rules, these systems often lack transparency and reach their limits with regard to efficiency and maintainability. Since the 1990s the idea of an integrated product- and process planning and consequently of the central storage of product, process and resource information has been focused in digital manufacturing systems [5]. Thereby, product, process and resource data is interlinked in order to describe the assembly processes thoroughly in the phases of assembly planning and production.

The databases composing the backbones of the modern IT tools mentioned above can be utilized to support the assembly planning process. Depending on the nature of the new planning task, [6] distinguishes the repetitive, adaptive, variant or new planning approach. Repetitive planning is applied in the case a former process plan for the exact same part and manufacturing process already exists. Adaptive planning is used, if a similar assembly process plan for the same part is available, but the planning premises have been different at the time. In the case of production-technical similarity, variant planning can be applied. Completely new products imply the generation of a novel process plan [6]. Especially in the case of adaptive and variant planning data mining approaches are potentially useful to identify similar structures in product and process data, e.g. in order to form a part family [7, 8].

2.2. Reduction of product and process complexity by data mining application

Data mining describes the concept of applying data analysis and discovery algorithms to produce a particular enumeration of patterns or models over data [9]. In general, predictive and descriptive data mining tasks can be distinguished. While predictive data mining is applied to find relationships between a dependent (target) variable and the independent variables in the dataset in order to make predictions on new and unlabeled data, descriptive data mining serves to produce understandable and useful patterns describing a complex dataset, yet without any prior knowledge of what patterns exist [10]. To reuse existing knowledge for future product generations and to reduce product and process complexity in assembly planning, classification and clustering as representatives of predictive resp. descriptive data mining methods can be used.

In terms of classification, the input dataset consists of n examples $x=(x_1,...,x_n)$ with each i – th example representing a vector of observed variables $x_i=(x_{i1},...,x_{im})$ of the total set of independent variables $X=(X_1,...,X_m)$ and having a class label g_i [11]. The input dataset serves as the basis for the learning of the function $f(x)=\hat{g}$, which describes the relationship between the input variables x and their correspondent class label g. $\hat{g}$ represents the target value predicted by the learning function. The extracted relationships between attribute values and class labels are used to assign labels to previously unseen data.

A wide spread classification method is the naive Bayes classifier. Hereby, the learning function f is interpreted as a classifier which predicts the class of a given observation x_i based on the probability $p(g_i)$ of the class and on the likelihood $p(x_{i1},...,x_{im}\mid g_i)$ of the feature values given the class g_i [12]. It assumes independency of the attributes $p(x_{i1},...,x_{im})$:

$$p(g_i\mid x_{i1},...,x_{im})=p(g_i)\prod_{j=1}^{m}p(x_{ij}\mid g_i) \tag{1}$$

In contrast, clustering is a typical approach for the reduction of data complexity by dividing a population of n examples $x=(x_1,...,x_n)$ into smaller subpopulations (clusters), so that the pairwise dissimilarities between the examples in one cluster tend to be smaller than those in different clusters [11].

The k -means clustering algorithm optimizes this criterion and is one of the most popular iterative clustering methods [11]. The parameter k representing the number of clusters must be specified in advance [12]. Being applied to datasets with quantitative variables, the algorithm begins to choose k random data points as cluster centroids. Subsequently, all examples of the input dataset $(x_1,...,x_n)$ are assigned to their closest cluster center using the Euclidean distance as dissimilarity measure. Then the centroid (mean)

$$\mu_k=\sum_{i\in S_k}\frac{x_i}{n_k} \tag{2}$$

of each newly formed cluster S_k is calculated, with n_k being the number of points in S_k. The centroids are set to be the new cluster centers. This process of minimizing the sum of squared errors [13]

$$J_K=\sum_{k=1}^{K}\sum_{i\in S_k}(x_i-\mu_k)^2 \tag{3}$$

is executed iteratively until the same examples of the input dataset are assigned to the same corresponding cluster in consecutive rounds.

Since assembly planning faces the challenge of high variance, cluster building techniques can be used to identify similar products or assembly processes and to group them according to the similarity of their characteristics. Beyond that, classification mechanisms are required to classify new and unseen assembly structures into the identified clusters in order to make use of the discovered correlations. Accordingly, adaptive and variant planning are promising use cases for data mining application in the assembly planning context. To realize and assess the potential of the data mining-supported assembly process planning the entire process of knowledge discovery in databases must be proceeded.

2.3. Knowledge discovery in databases

KDD describes the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [14]. The procedure of KDD is described in different process models and has been developed for both scientific and practical contexts. These models differ by the number, names and presentation of their iteratively performed steps, but describe substantially similar procedures of knowledge discovery. One of the leading models with high practical relevance is the Cross-Industry Standard Process for Data Mining (CRISP-DM) consisting of six major steps [15]: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment (Figure 1).

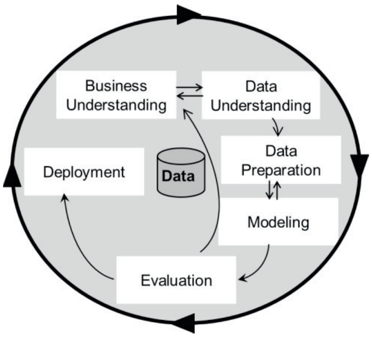

Fig. 1. CRISP-DM process steps [15].

Within Business Understanding the description of the knowledge acquisition objectives and the definition of data mining tasks (predictive or descriptive) are focused. The existence, availability, possibilities of access and quality of relevant data are reviewed during Data Understanding. The next step (Data Preparation) comprises the often time- and personnel-consuming transformation of data to a data mining method-suitable format, e.g. by selecting specific datasets and variables or by removing noisy data. The application of specific data mining techniques, e.g. classification or clustering, to extract patterns from data is performed in the Modeling step. Results are assessed and interpreted by experts of the future application area within the Evaluation step. Assuming positive and valid results the developed model can be deployed in the considered field and consequently be used e.g. for planning- and decision-making support.

3. Need for research

The KDD approach is based on the assumption that large and merged homogeneous databases exist, which is not always fulfilled in the production context. Here both the integration of heterogeneous sources, such as the PDM or digital manufacturing systems, and the harmonization of various data formats have to be ensured beforehand.

In contrast to the production phase, where a continuous data collection can be achieved due to the application of sensors and manufacturing execution systems, the assembly planning phase bases on manually generated data with smaller volume. Nevertheless, the data contains valuable knowledge. At present its utilization in order to support the assembly process planning procedure is limited [16]: The prevailing method to draft an assembly process plan reverts in the first instance to expert know-how to select a resembling one from the population of existing plans. The selected plan is adapted to the latest planning premises in the graphical user interface of the digital manufacturing system as data source and regular working environment of the process planner in the next step.

The current demand for the improved support arises due to the increasing planning complexity and the skill-intensive selection of existing assembly process plans. To meet the relevant challenges of assembly process planning as well as to manage and assess planning knowledge with decreased efforts, data mining techniques can be utilized in KDD process.

4. Concept of data mining-supported generation of assembly process plans

The process planning environment in the manufacturing industry is characterized by an increased demand for intelligent solutions to assist process planners in the efficient design of assembly process plans [17]. For this purpose, data from the PDM and digital manufacturing system can be analyzed in order to identify hidden and yet unknown patterns in the parameters that determine assembly processes. These patterns represent a basis for the creation of suitable assembly process templates for future product generations.

The formation of these templates is set up in four main processes regarding Data Preparation and Modeling: First product data, initially represented by the bill of materials and 3D shape metadata, is aggregated to subassembly specific feature vectors. These are used to train a classification model which is able to differentiate between the various subassembly types automatically. Once being differentiated, a clustering algorithm is applied to all variants of each subassembly type in order to find similar subassembly groups within one type and to reduce variant complexity. The clustering of process data is performed additionally and independently from the product data. Finally, a function is trained to map the product clusters to their respective process clusters.

4.1. Classification of subassembly types

The final product is composed of a number of subassembly types, with each of them providing a specific function. Regarding e.g. engine assembly, cylinder head, crankcase and electric generator are just a few examples for subassembly types. In order to generate reasonable clusters, a classification model is required to distinguish between the various functional subassembly types $T = 1,...,t$.

Consequently, a preliminary focus is set on establishing feature vectors to characterize and differentiate the subassembly types. Features vectors are required to be independent from specific part identification numbers and contain information such as outer dimensions, weight, center of gravity and amount of both designed and standard components of the entire subassembly. Furthermore, the distribution of outer dimensions and weights for the single components of one subassembly is represented by the top n of the largest and heaviest parts.

In the training phase, the subassembly type is added manually to product data and serves as target variable for the classification model. The exemplarily input dataset for product data is shown in Table 1.

Table 1. Exemplary input dataset for product data

Sub-assembly ID	Sub-assembly type	Dimension X Y Z	Weight	Center of Gravity X Y Z	Part 1	Part 2	..
ID_1	u						
ID_2	v						
…							
ID_n	w						

4.2. Clustering of same type subassemblies

Each subassembly type is prone to variant multiplicity. Having differentiated various subassembly types, complexity of subassemblies is reduced by clustering them according to the similarity of their characteristics with agglomerative and k –means clustering. Prior to this, all attribute values are normalized to the interval [0, 1] to achieve equal attribute weights during the clustering.

As a result, a partition $S_t = \{S_1, S_2, ..., S_{k_t}\}$ of the collectivity of subassemblies of the same type t is created, where a single cluster S_i of subassemblies is referred to as *product cluster* in the following text. k_t is the number of clusters determined individually for each subassembly type t with the help of a supplemental agglomerative clustering algorithm. These clusters that resemble the most are merged consecutively until a single big cluster remains [11]. A dendrogram is used to visualize the tree structure arising during the agglomerative clustering process and to determine a suitable number of clusters.

The assumption of the data mining model is that subassemblies being sufficiently similar concerning their characteristics also require similar assembly processes. Hence, clustering of assembly processes is realized independently from product clustering for each subassembly type t . The input data for the process clustering consists of existing assembly process plans and is composed of both textual process descriptions and detailed predetermined motion time system (PMTS) analyses. Thereby, each process description is broken down to its basic movements and the total number of basic movements is summed up for each subassembly. The resulting table shown exemplarily in Table 2 serves as input for the application of the k -means clustering algorithm, which is executed on the normalized frequencies of PMTS codes.

Table 2. Exemplary input dataset for process data

Sub-assembly ID	Sub-assembly type	Process Description	Process Execution Time	Time Study ID	PMTS Code 1	PMTS Code 2	PMTS Code 3	…
ID_1	Type u							
ID_2	Type v							
…								
ID_n	Type w							

The process clustering results in a partition $R_t = \{R_1, R_2, ..., R_{k'_t}\}$ for the collectivity of assembly processes required to assemble subassembly type t with k'_t being the number of process clusters determined individually for assembly type t . A cluster R_i of assembly processes is referred to as *process cluster*.

4.3. Product-process-mapping

Due to the existing single linkages between subassemblies and their required assembly processes, which are available in the digital manufacturing system, a mapping function is derived to assign product clusters containing subassemblies with similar characteristics to their respective process clusters. This is done separately for each subassembly type t and is illustrated exemplarily in Table 3.

Table 3. Exemplary input dataset for the computing of the product-process mapping for assembly type t

Subassembly ID	Product Cluster	Process Cluster
ID_1	S_1	R_2
ID_2	S_4	R_3
…	…	…
ID_n	S_3	R_3

The generation of the mapping function is done by means of a naive Bayes classifier, where the process cluster is set to be the target variable (product-process-mapping).

In future application as an assistance function, a subassembly of a future product generation is classified into an existing product cluster. Based on the product cluster, product-process-mapping deduces the process cluster with the highest probability. With the help of the process cluster, a suitable assembly process template can be generated.

5. Case study in the automotive industry

The described concept for the data mining-supported generation of assembly process plans was applied to multi-variant subassembly types of the automotive industry according to the CRISP-DM process steps. Product data of three subassembly types (13000 datasets, 24 attributes) and the corresponding assembly work plans describing manual assembly tasks for different product variants in series production (700 datasets, 200 attributes) were analyzed with RapidMiner 5.3, an open source software tool for data mining and machine learning [18].

5.1. Data preparation

The clustering data is extracted from the PDM and the digital manufacturing system. Since both product and process data are organized in hierarchical structures, the transformation into a flat data table is the first step to be performed.

In order to provide subassembly features, outer dimensions, weight and center of gravity for the position of installation are extracted from the 3D shape representation. Additional information describing the bill of materials content is gathered from the various hierarchical levels of the bill of materials and stored in a flat example set (Figure 2).

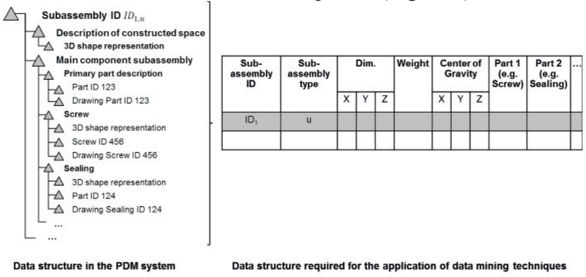

Fig. 2. Transformation of product data into a flat table.

A similar transformation step is realized concerning the assembly process data from the digital manufacturing system, which additionally is enriched with the detailed itemization of process steps from PTMS.

5.2. Modeling

Subsequent to the generation of feature vectors, value differences in the feature attributes for different subassembly types are analyzed. Exemplarily, this is shown as a deviation plot in Figure 3.

The lines depict mean attribute values for each subassembly type. Deviations are added in order to demonstrate the attribute value ranges. Based on this input data, a classification model is trained, which is able to extract and formalize the differences in the attribute values and to use this knowledge to classify new and unseen subassembly types. The model is trained on one product generation and applied to a second one in order to validate the approach. The obtained overall accuracy is 97.83% on three different subassembly types.

In the next step, a k -means clustering algorithm is performed separately for each subassembly type and the results are checked for plausibility. They are satisfying in the way that the applied algorithm differentiated the product data in valid subject-specific groups, e.g. concerning the distinct country or operation editions of the subassembly.

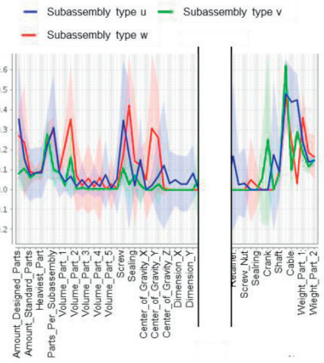

Fig. 4. Visualization of feature vectors.

Data for process cluster formation is extracted from a digital manufacturing system listing the basic movements of the diverse assembly processes that each variant of one subassembly type requires.

After accomplishing product and process clustering, the subassembly ID is extracted together with the affiliation to the respective cluster separately for each subassembly type (Table 3). To derive the most probable assembly process cluster for the subassembly type, a naive Bayes classifier is trained to assign product clusters to the process cluster with the highest probability. A resulting product-process-mapping for one subassembly type is presented in Figure 5.

The figure indicates the matching probability of a process cluster given a certain product cluster. In the cases of product clusters 0, 1, 2, 6 and 7, the process clusters 0, 1 resp. 3 can be deduced with maximum likelihood. As to product clusters 3 and 4, the mapping is not entirely decisive, yet process cluster 0 is a lot more probable than process clusters 2 resp. 1. Concerning product cluster 5, no process cluster could be mapped. Since productive data is used, the work plans for series production of these subassemblies have not yet been released.

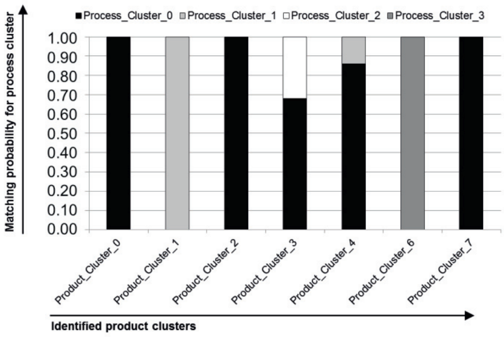

Fig. 5. Product-process-mapping.

5.3. Evaluation

The product-process-mapping probabilities show a comprehensive connection between product and process clusters. The assumption that subassemblies being sufficiently similar concerning their characteristics also require similar assembly processes can consequently be affirmed. Data mining techniques allow the revelation of these connections and permit the application to future product generations.

5.4. Deployment

Once the model generation is completed successfully, it can be applied to future product generations. After development of a new subassembly, the bill of materials including the 3D shape representations is uploaded to the PDM system and represents the starting point for assembly process planning. Correspondingly, the process planner uses this data as input for the model application, which selects the product cluster that fits the input data with the highest similarity and probability and deduces the process cluster with maximum likelihood. The process cluster serves as basis for the following generation of the process template, which can be transferred directly to the current digital planning project. The process planner reviews the proposed template and adapts it to the latest premises of the planning context. Thereafter the adjusted new planning data is added to the analysis datasets for the recalculation of product and process clusters. In this way, the mapping model will be refined with every iteration loop.

6. Conclusion and further developments

The developed approach for the data mining-supported generation of assembly process plans provides a new assistant function in the field of digital assembly process planning. The former routine work of identifying a similar and suitable process plan that has to be adapted manually is primarily based on expert knowledge. Due to the fact that the presented approach is based on planning data compiled during the entire product emergence process, this routine becomes more transparent and leads to a partial automation of the assembly planning process as well as a faster and easier attainment of a comparable planning level. In this way a first assembly process plan can be established at a very early stage of the product emergence process.

In future the detailed configuration of templates should be focused and the applicability to different production and assembly domains needs to be investigated. The integration of the assistance function in the digital manufacturing system has to be ensured in order to provide a fully integrated data mining-supported assembly process planning function.

Acknowledgements

This paper represents the background, objectives and first results of the research project “Prospective Determination of assembly work content in Digital Factory (Pro Mondi)”. This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the Framework Concept “Research for Tomorrow’s Production” (funding number 02PJ1110) and managed by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the contents of this publication.

References

[1] Bley H, Zenner C. Variant-oriented Assembly Plannning. CIRP Annals – Manufacturing Technology 2006;55:23-28.

[2] Eigner M, Stelzer R. Product Lifecycle Management – Ein Leitfaden für Product Development und Lifecycle Management. Berlin: Springer; 2009.

[3] Thaler K. Regelbasiertes Verfahren zur Montageablaufplanung in der Serienfertigung. Berlin: Springer, 1993.

[4] Dong T, Tong R, Zhang L, Dong J. A knowledge-based approach to assembly sequence planning. Int J Adv Manuf Technol 2007;32:1232-1244.

[5] VDI 4499 Digital Factory - Fundamentals. Duesseldorf: Verein Deutscher Ingenieure, 2009.

[6] Eversheim W. Organisation in der Produktionstechnik – Arbeitsvorbereitung. Berlin: Springer; 2002.

[7] Deuse J. Fertigungsfamilienbildung mit feature-basierten Produktmodelldaten. Aachen: Shaker; 1998.

[8] Eversheim W, Deuse J. Formation of Part Families based on Product Model Data. Production Engineering 1997; IV/2: 97-100.

[9] Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases: AI Magazine 1996;17:37-54.

[10] Wang K. Applying data mining to manufacturing: the nature and implications. J Intell Manuf 2007;18:457-495.

[11] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning - Data Mining, Inference, and Prediction. New York: Springer; 2009.

[12] Witten IA, Frank E. Data Mining – Practical Machine Learning Tools and Techniques. San Francisco: Morgan Kaufmann; 2005.

[13] Ding C, He X. Cluster Structure of K-means Clustering via Principal Component Analysis. In: Dai H, Srikant R, Zhang C, editors. Advances in Knowledge Discovery and Data Mining, Proceedings of the 8th Pacific-Asia Conference PAKDD 2004. Berlin: Springer; 2004. p. 414-418.

[14] Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knnowledge Discovery: An Overview. In: Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R, editors. Advances in Knowledge Discovery and Data Mining. Mensolo Park: AAAI Press 1996, p. 1-30.

[15] Chapman P, Clinton J, Kerber R, Khabaza T, Reinhartz T, Shearer C, Wirth R. CRISP-DM 1.0 – Step-by-step data mining guide. SPSS; 2000.

[16] Erohin O, Kuhlang P, Schallow J, Deuse J. Intelligent Utilisation of Digital Databases for Assembly Time Determination in Early Phases of Product Emergence. Procedia CIRP - 45th CIRP Conference on Manufacturing Systems 2012;3:424-9.

[17] Wallis R, Erohin O, Stromberger F, Deuse J. Data Mining Application in Digital Process Planning – RapidMiner in Automotive Industry. In: Fischer S, Mierswa I, Moreira JM, Soares C, editors. Proceedings of the 4th RapidMiner Community Meeting and Conference. Aachen: Shaker-Verlag; 2013, p. 165-174.

[18] Mierswa I, Wurst M, Klinkenberg R, Scholz M, Euler T. YALE: Rapid Prototyping for Complex Data Mining Tasks. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2006. New York: ACM Press, p. 935-940